Partitioning the target space in multi-output learning

Isaac Triguero Isaac.Triguero@nottingham.ac.uk
University of Nottingham, School of Computer Science, Nottingham, United Kingdom

Marcio Basgalupp basgalupp@unifesp.br
Federal University of São Paulo, Institute of Science and Technology, São Paulo, Brazil

Ricardo Cerri cerri@dc.ufscar.br
Federal University of São Carlos, Department of Computer Science, São Carlos, Brazil

Leander Schietgat Leander.Schietgat@cs.kuleuven.be
KU Leuven, Department of Computer Science, Leuven, Belgium

Celine Vens Celine.Vens@kuleuven-kulak.be
KU Leuven Kulak, Department of Public Health and Primary Care, Kortrijk, Belgium

Keywords: multi-label classification, multi-target regression

Abstract

Multi-output prediction problems are solved by either a global or a local prediction strategy. In this study, we investigate whether there exist strategies in between these two extremes that result in a better predictive performance. Preliminary results suggest that, while in general the global approach is best, in some cases it can be outperformed by a different strategy that uses a small number of target subsets.

1. Introduction

Traditional prediction problems deal with a set of instances that have a single target value associated with them. This target attribute can be nominal (classification problem) or numeric (regression problem). Several real life problems, however, have a *set* of target attributes: instead of a single property, one is interested in predicting multiple properties. This setting is known as multi-target prediction. Applications include the prediction of river water quality parameters from bioindicator data or the prediction of the number of employees across many employment types for a specific metropolitan area (Spyromitros-Xioufis et al., 2016). A very related setting is multi-label learning, where multiple class labels have to be predicted for the examples. A multi-label prediction problem can be viewed as a binary multi-target regression problem (Tsoumakas et al., 2014). Applications include protein function prediction, document annotation, etc.

Generally there are two strategies to tackle multi-output problems. In the first strategy, which we call the local strategy, one converts the problem into a set of single-output problems, and applies a standard prediction model. In the global strategy, one keeps the multi-output structure and applies a multi-output prediction model. Several machine learning techniques have been extended to multi-output problems: decision trees, support vector machines, artificial neural networks,... Generally, global approaches are preferred in terms of efficiency and model size, while local approaches benefit from their simplicity. In terms of predictive performance, global strategies are better than or equal to local strategies (Kocev et al., 2013).

Although several studies that exploit target or label dependencies in local approaches have been proposed, e.g. classifier chaining methods (Read et al., 2011; Spyromitros-Xioufis et al., 2016), these methods still predict a single output at once. To our knowledge, there exist no studies that apply a strategy other than global or local learning. However, because of these dependencies that may exist between targets, it may well be that grouping targets to be learned together per-

Appearing in *Proceedings of Benelearn 2016*.

forms better than the local or global approach. This is exactly what we investigate in this study: is there a partition of targets that, when each subset in this partition is treated as a separate global prediction problem, outperforms these two extremes?

2. Method

The number of partitions of a set (also known as the Bell number) grows very quickly. For instance, a set of 10 elements has 115,975 possible partitions; a set of 16 elements has 10,480,142,147. However, noting that a set of n elements has 2^n possible subsets, this means that the 10,480,142,147 partitions are composed of 65,536 subsets only, which makes it possible to conduct an exhaustive search over all partitions. More precisely, we propose to train a multi-output classifier on each of the 2^n possible subsets of targets and store the predictions made by this classifier on a hold-out validation set. Then, we perform an efficient generation of all set partitions using the procedure outlined in (Orlov, 2002), and for each partition combine the corresponding predictions. The partition that returns the best predictive performance over the validation set is returned and used to obtain predictions for the final test set. The multi-output classifier that we used constructs multi-target predictive clustering trees (Kocev et al., 2013).

While most multi-target regression benchmark datasets typically have a limited number of targets, in the multi-label classification case, the situation is different. The typical application domains there have a large label space, often containing a few hundred labels, which renders the exhaustive approach infeasible. For this setting, we devised a genetic algorithm approach that evolves a population of partitions in order to find a (local) optimum.

3. Results

Here, we show results for the multi-target regression task, where we studied the optimal partition of targets returned by the exhaustive search strategy. We used 7 benchmark datasets, available at `http://mulan.sourceforge.net/datasets-mtr.html`. The main research question addressed is whether there exist better strategies in between the extremes of global and local classification. Table 1 shows the results: it lists the number of targets and the number of subsets in the optimal partition. A number of subsets equal to one corresponds to the global multi-target approach, while a number of subsets equal to the number of targets corresponds to the local approach. We observe a large discrepancy between optimising predictive performance on the training set or on a hold-out validation set. While the former leans towards the local approach, the latter seems to promote the global approach. These results seem to suggest that the local approach may be more sensitive to overfitting. An exception seems to be the Water Quality dataset (WQ), where in the training set optimisation, a partition with only two subsets is returned. The targets here correspond to the abundance of 14 organisms that are found in river water samples. Interestingly, the two subsets distinguish the plant and animal organisms. In the validation set optimisation, three subsets appear. While the organisms of the same biological order remain together, two of the subsets contain a mix of plants and animals. It remains to be investigated whether this makes sense biologically. W.r.t. our original research question, focussing on the results obtained on the validation set, we can conclude that generally the global approach is best. In some cases the global and local approaches can be outperformed by a different partition; however, this partition always contains a small number of subsets.

Table 1. Exhaustive search results. The table shows for each dataset the number of targets and the number of subsets in the optimal partition. Results are obtained by minimising mean squared error on the training set and on a hold-out validation set.

Data set	Nb Targets	Nb Subsets	
		Training	Validation
oes10	16	16	1
oes97	16	16	2
rf1	8	4	1
rf2	8	4	1
scm1d	16	16	1
scm20d	16	16	1
wq	14	2	3

4. Conclusion

This study conducts an exhaustive approach to obtain an optimal partitioning of the outputs in multi-output learning. For a large number of outputs (e.g., in multi-label classification tasks) a genetic algorithm can be used to evolve a (sub)optimal partition. Preliminary results on multi-target regression show that a global prediction strategy generally gives the best predictive performance. These results differ from those obtained by (Jacob et al., 2009) for the related task of multi-task learning, where learning (linear) functions in partitions always improves upon the global approach.

References

Jacob, L., philippe Vert, J., & Bach, F. R. (2009). Clustered multi-task learning: A convex formulation. In D. Koller, D. Schuurmans, Y. Bengio and L. Bottou (Eds.), *Advances in neural information processing systems 21*, 745–752. Curran Associates, Inc.

Kocev, D., Vens, C., Struyf, J., & Deroski, S. (2013). Tree ensembles for predicting structured outputs. *Pattern Recognition*, *46*, 817–833.

Orlov, M. (2002). Efficient generation of set partitions.

Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2011). Classifier chains for multi-label classification. *Machine Learning*, *85*, 333–359.

Spyromitros-Xioufis, E., Tsoumakas, G., Groves, W., & Vlahavas, I. (2016). Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, 1–44.

Tsoumakas, G., Spyromitros-Xioufis, E., & Vlahavas, I. (2014). Drawing parallels between multi-label classification and multi-target regression. *ECML PKDD 2014 Workshop on Multi-Target Prediction*. Nancy, France.